GR∇DIENT0

Secure Machine Learning on sensitive data

With the DQ0 data quarantine

2019

www.gradient0.com

Data and Machine Learning

The rapid rise of artificial intelligence, and in particular its most important discipline, machine learning, is based on three factors: scientific breakthroughs in computer science, sufficient computing power and data. Lots of data. Data is the foundation of the AI renaissance, which is much more than a trend. Data-driven processes are so successful that they will have a decisive influence on progress across industries and will have a say in the long term. Using data sensibly and on a large scale means strengthening growth and profitability and opening up new opportunities that would not be possible without AI and without the data.

Though, two questions arise:

- ▽ How can I provide my data for machine learning?
- ▽ How can I guarantee that I will stay the owner of my data and that the processing of the data is secure and fully under control?

In order to develop or implement data-driven processes, the help of experts from the field of data science, specially trained computer scientists or mathematicians who are skilled in advanced statistical analysis and methods of machine learning, is required. As a business owner, I have two options: either I establish one or more data science teams within my organization, or I get these services through partners. For both ways, I need a secure interface that defines what data is provided to the internal or external data science team.

The data in question are often those that contain information worth protecting.
Data protection finally gets the attention it deserves, not least because of the EU data protection directive and national regulations. "Data ownership" is an important issue. I have to make sure that I have control over my data at all times. Personal data or company secrets may never be disclosed (without explicit consent).

AI is growing up

Machine learning is no longer a new discipline. Its tools and methods are becoming increasingly mature and accessible to a broader group of professionals. At the same time, business processes are being established around the concept of data marketplaces, where companies can quickly and easily make their data available to data scientists in order to benefit themselves from new solutions and optimized processes or to directly realize the value of their data.

Artificial intelligence is everywhere. Although their foundational computer science research is still at the very beginning, their applications and tools for specific domains are already very mature. Often, general, pre-trained machine learning models are being used; in the case of speech processing, for example, models pre-trained on large and general vocabularies can be "retrained" on the actual data, that is, specialized for the particular application. But what is the interface for this process?

The standard way today: The data is brought to where it can be processed by data scientists. To ensure privacy, the data is usually pseudonymized (replacement of personal entities), generalized or perturbated (generalized or altered in a way that individual records are no longer identifiable) or minimized (removed from allegedly unused properties of the data).

This approach is not a solution. The anonymization or pseudonymization process is complex and domain-specific, there is no general solution for this. For some data (for example certain texts or images), anonymization is hardly an option. In addition, this process is error prone and not sufficiently secure. In numerous publications and competitions secret information was disclosed from supposedly sufficiently anonymized datasets. There is no method to algorithmically determine whether data itself has actually been successfully anonymized.

And then: how? or rather where? are the data pseudonymized anyway? Here, again, we need a secure gateway to the data to be able to process them – a chicken or egg dilemma.

Additionally, this approach does not solve the problem of the interface of the machine learning toolsets. The data science team in place needs to provide or integrate those themselves.

DQ0

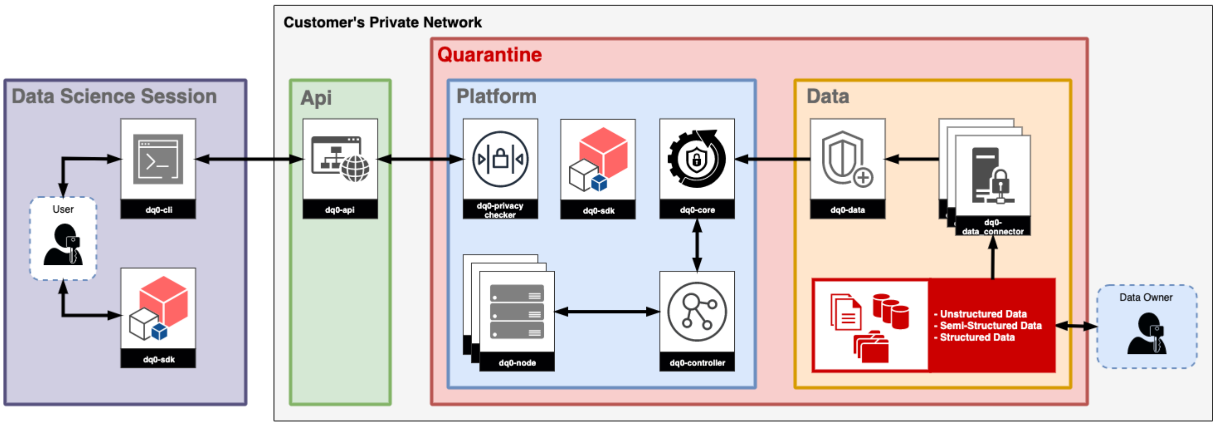
The DQ0 data quarantine solves both problems. It offers a safe haven for the data and at the same time makes it accessible for analysis in such a way that a data science team can work directly with the data with its usual tools.

DQ0 turns the tables: the data does not move to the analysis; the analysis moves to the data.

DQ0 quarantines the existing data. This ensures that the data is protected against unauthorized access. The data will not be changed or released at any time.

DQ0 implements the privacy policy at the time of the request to the data. The fundamental principle here is that of differential privacy, which Gradient Zero extended to the area of machine learning. Any response that leaves the quarantine via the secure DQ0 API will be checked for strict security and privacy compliance.

The DQ0 API thus provides a secure interface for performing machine learning on the data. The data science team can iteratively develop AI models without exposing sensitive information about the data.



As seen in the diagram above, DQ0 is completely installed within the customer's network. In the optional cloud variant of DQ0, the network is also set up in a way that only the customer has access and control rights. Only the port for the DQ0 API remains open to the public. Every response that leaves the quarantine is checked beforehand by the "DQ0 Privacy Checker".

The data owner has complete data sovereignty. DQ0 can handle any type of data: structured data as well as text documents, images or any other format.

Access to the data is strictly read-only, implemented by the DQ0 data connectors and the DQ0 data shield.

Users who want to develop models on the data log in via the DQ0 API and use the DQ0 Software Development Kit (SDK) to define and test the models. The models can then be sent to the DQ0 customer instance via the API and the DQ0 Command Line Client. There they are trained on the real data. The return value is a response checked by the DQ0 Privacy Checker comprising the quality of the current model version.

The DQ0 API also enables general, aggregated information and schemata of the data to be queried, as well as the generation of synthetic data, in order to facilitate the necessary data preprocessing and feature engineering steps.

Standard models can be easily transferred to the DQ0 instance via the intelligent DQ0 interface in order to quickly test their suitability for the specific problem.

With the optional studi0 integration, which is available in the cloud version of DQ0, a comprehensive data science platform for workflow management, data and model versioning and high-parallel model fitting can be used.

Differential Privacy and Machine Learning

DQ0 consistently hides the customer's data in a secure, internal network. Since the models move to the data and not vice versa, no information about the data has to be released at any time. But how does DQ0 ensure that the answers that external data scientists receive are adequately protected and at the same time useful for their work?

> DQ0 relies on the principle of differential privacy, which is used consistently in the quarantine in all phases of machine learning development.

Differential Privacy (DP) is a definition that mathematically guarantees that anyone who sees the result of a differential privacy analysis will draw the same conclusion about a person's private information, regardless of whether that person's private information was included in the analysis or not.

To put it differently: Queries to a data record do not differ according to differential privacy (or only to a definable, limited extent) if a data point is added to the data record or if it is removed from it. Therefore, in DP-compliant queries, no conclusions can be drawn about individual data points (e.g. people).

Differential Privacy guarantees protection against:

- Membership disclosure (e.g. does a person have a specific account?)
- Attribute disclosure (the disclosure of an attribute of one data entry)
- Identity disclosure (the clear assignment of an entry to a person)

Differential Privacy is mathematically defined as: A randomized computation *M* satisfies *ε*-differential privacy, if for any adjacent data sets *x* and *x*' and any subset *C* of possible outcomes Range *(M)*, holds:

$$Pr[M(x) \in C] \leq \exp(\epsilon) \times Pr[M(x') \in C]$$

This means, A randomized algorithm M provides sufficient privacy protection if the ratio between the probabilities that two *adjacent* datasets give the same answer is bounded by exp(ε).

However, the protection of information according to this principle requires special care in the area of machine learning. That's because with extended statistical methods, the ε-differential privacy property could be exploited in order to obtain conclusions about individual data points via a sufficient number of queries.

A possible attack could be a so-called "shadow model", which in turn evaluates the differential privacy compliant responses in order to obtain information worthy of protection.

In its simplest form this attack leverages the „overfitting gap“ of a model to gain a membership disclosing privacy breach. This difference between the error curves of training and test is greater the more specific the model was trained for the test data. Because machine learning models achieve their performance by adapting to presented training data so that they minimize the sum of the errors in the differences of the predicted values and the target values, thus, "learning" from the the training data, it can easily happen that they, with enough degrees of freedom, simply memorize the data. The predictive performance on the presented training data is then very high, but low on new test data.

The attack mentioned makes use of this property: it looks at the confidence in the prediction of a presented data point (i.e. how certain is the model that the data set falls, for example, in a predicted category) and concludes from a high confidence that this data point was part of the training set and

therefore must have been included in the allegedly hidden original data set.

The DQ0 Model Checker counter-acts this attack by automatically checking each model that is to be executed within the quarantine. Only if this test is passed according to the implemented differential privacy criteria can the model be used from outside via the DQ0 API beyond actual training.

A prediction request on the model within the domain of the customer is possible at any time.

The DQ0 Data Quarantine ensures at all times that the API will only release information about the data that will never lead to privacy breaches, even with advanced statistical procedures.

It relies on innovative methods for data security for machine learning.

The method has been scientifically verified; the implementation is being certified by TÜV Austria.

Tl;dr

Data is valuable. They can form the basis for the optimization of existing processes, the development of new insights or possibilities and for lasting economic success. However, this can only succeed in a sustainable way if the security of data becomes a top priority. Gradient Zero developed the DQ0 data quarantine for this purpose - by data scientists for data scientists, built without compromise for data security. With DQ0, internal or external data science teams gain access to sensitive data without ever putting sensitive information at risk.

DQ0 is the secure interface between data and machine learning. Encryption and data-secure queries are not simply enforced by DQ0 through defined policies, rather, they are algorithmically baked into the DQ0 software. The DQ0 SDK offers a standardized interface for generally available models as well as for self-developed models.

Whether sensitive medical data, user information with personal data or secret machine data, DQ0 is the simple and secure interface to the data-driven future.

About Gradient Zero

Gradient Zero is a leading machine learning company based in Vienna, Austria. We offer cross-industry machine learning solutions.

Artificial intelligence has become an established business and research area. Successful AI projects, however, require well-designed solutions with carefully selected techniques and a thorough understanding of the underlying methods, their implementation and data security requirements. AI is here to make things easier, faster, and more convenient - that's our mission.

Contact us

e-mail:	office@gradient0.com
phone:	+43 660 4259199
web:	www.gradient0.com